From *SIAM News*, Volume 42, Number 5, June 2009

CSE 2009

Theory, Algorithms, Applications: Advances in Model Reduction

By Christopher Beattie and Serkan Gugercin

Dynamical systems theory provides some of the principal tools used in the modeling, prediction, and control of physical phenomena as diverse as signal propagation in the nervous system, storm surges before an advancing hurricane, temperature control and sterilization in food processing, and microchip design. Direct numerical simulation of mathematical models has been one of very few available means for studying these and other extremely complex systems. The need for greater accuracy leads to the inclusion of additional detail in the models, which often need to be coupled to models of other complex systems that can operate in different time and spatial scales. The resulting computational burden can be overwhelming, making unmanageably large demands on resources. Efficient model utilization becomes a necessary component of simulations in such large-scale settings. This is the main motivation for model reduction.

In the broadest sense, any applied mathematician who works with models of real-world phenomena is engaged in "model reduction" of one sort or another. We use the term in the sense that has become prevalent—the use of systems-theoretic techniques to create smaller and cheaper models that carefully encode fine-scale dynamical features of the original system, ultimately allowing close mimicry of the input/output map.

Although specific settings vary, a dynamical system can be described in terms of an input/output map $\mathcal{S}$: $L^2([0,\infty),\mathbb{R}^m) \to L^2([0,\infty),\mathbb{R}^p)$ with the state-space representation

$$\mathcal{S}\!: \mathbf{u} \mapsto \boldsymbol{\Theta}$$
$$\begin{cases} \mathbf{E}\dot{\mathbf{y}}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{f}(\mathbf{y}(t),\mathbf{u}(t)), \\ \boldsymbol{\Theta}(t) = \mathbf{g}(\mathbf{y}(t),\mathbf{u}(t)), \\ \text{with } \mathbf{y}(0) = 0, \end{cases}$$

where $\mathbf{E}, \mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{f}: \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$, and $\mathbf{g}: \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^p$. The vector $\mathbf{y}(t) \in \mathbb{R}^n$ describes the internal *state* of the system; $\mathbf{u}(t) \in \mathbb{R}^m$ is the input (excitation); and $\boldsymbol{\Theta}(t) \in \mathbb{R}^p$ is the output (measurement).

The internal states of the original system model often evolve along trajectories that do not fully occupy the state space, but hew fairly closely to subspaces or manifolds of substantially lower dimension—that is, the input/output map behaves almost as if it had far fewer internal degrees of freedom. The goal of model reduction is to discover and mimic lower-dimensional dynamical systems of this type, creating an input/output response as close to the original as possible. The resulting reduced model, $\mathcal{S}_r$, is described analogously:

$$\mathcal{S}_r\!: \mathbf{u} \mapsto \boldsymbol{\Theta}_r$$
$$\begin{cases} \mathbf{E}_r\dot{\mathbf{y}}_r(t) = \mathbf{A}_r\mathbf{y}_r(t) + \mathbf{f}_r(\mathbf{y}_r(t),\mathbf{u}(t)), \\ \boldsymbol{\Theta}_r(t) = \mathbf{g}_r(\mathbf{y}_r(t),\mathbf{u}(t)), \\ \text{with } \mathbf{y}_r(0) = 0, \end{cases} \quad (1)$$

where $\mathbf{E}_r, \mathbf{A}_r \in \mathbb{R}^{r\times r}$, $\mathbf{f}_r: \mathbb{R}^r \times \mathbb{R}^m \mapsto \mathbb{R}^r$, $\mathbf{g}_r: \mathbb{R}^r \times \mathbb{R}^m \mapsto \mathbb{R}^p$, and the reduced system is of order $r \ll n$.

Successful model reduction methods should achieve the following goals:

- The reduced input/output map $\mathcal{S}_r$ should be uniformly close to $\mathcal{S}$ in an appropriate sense. That is, when presented with the same input $\mathbf{u}(t)$, the difference between full and reduced system outputs, $\boldsymbol{\Theta} - \boldsymbol{\Theta}_r$, should be *small* in a physically relevant norm over a wide range of system inputs, e.g., over all $\mathbf{u}$ in the unit ball of $L^2([0,\infty), \mathbb{R}^m)$.

- Strategies for obtaining $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{f}_r$, and $\mathbf{g}_r$ should lead to robust, numerically stable algorithms and should require minimal application-specific tuning with little or no expert intervention. It is important that model reduction methods be computationally efficient and reliable so that very large problems become and remain tractable; to allow the broadest level of flexibility and applicability in complex multiphysics settings, the methods should be robust and largely automatic.

Many speakers in the five sessions of the minisymposium on model reduction at this year's conference on CSE presented fresh developments, either theoretical or computational; others introduced exciting new application areas for model reduction in several science and engineering disciplines. The talks covered both linear and nonlinear dynamical systems.

In the case of *linear dynamical systems*, $\mathcal{S}$ is defined so that $\mathbf{f}(\mathbf{y},\mathbf{u}) = \mathbf{B}\mathbf{u}$, with $\mathbf{B} \in \mathbb{R}^{n\times m}$, and $\mathbf{g}(\mathbf{y},\mathbf{u}) = \mathbf{C}\mathbf{y} + \mathbf{D}\mathbf{u}$, with $\mathbf{C} \in \mathbb{R}^{p\times n}$ and $\mathbf{D} \in \mathbb{R}^{p\times m}$; $\mathcal{S}_r$ is defined analogously: $\mathbf{f}_r(\mathbf{y}_r,\mathbf{u}) = \mathbf{B}_r\mathbf{u}$, with $\mathbf{B}_r \in \mathbb{R}^{r\times m}$, and $\mathbf{g}_r(\mathbf{y}_r,\mathbf{u}) = \mathbf{C}_r\mathbf{y}_r + \mathbf{D}_r\mathbf{u}$, with $\mathbf{C}_r \in \mathbb{R}^{p\times r}$ and $\mathbf{D}_r \in \mathbb{R}^{p\times m}$. Laplace transformation of the input/output maps $\mathcal{S}$ and $\mathcal{S}_r$ yields (pointwise) multiplication operators in the transform domain by the *transfer functions* $\mathbf{H}(s)$ and $\mathbf{H}_r(s)$, respective

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$
and
$$\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r + \mathbf{D}_r\,.$$

This leads to two commonly used measures of the "closeness" of the systems $\mathcal{S}$ and $\mathcal{S}_r$:

$$\|\mathcal{S}-\mathcal{S}_r\|_{\mathcal{H}_\infty} = \max_{\omega\subset\mathbb{R}}\|\mathbf{H}(\jmath\omega)-\mathbf{H}_r(\jmath\omega)\|_2,$$
$$\|\mathcal{S}-\mathcal{S}_r\|_{\mathcal{H}_2} =$$
$$\sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty}\|\mathbf{H}(\jmath\omega)-\mathbf{H}_r(\jmath\omega)\|_{\mathrm{F}}^2 d\omega}.$$

Despite the extensive theory that has been developed for model reduction in the setting of linear systems, demands are often stringent and expectations accordingly very high; as a result, a wealth of interesting open problems remain even in this well-studied setting. For example, when the original system dimension n is very large, on the order of, say, hundreds of thousands or more, computing an effective reduced order model is a challenging but tractable problem and it is realistic to expect an *optimally close* reduced model for any arbitrary order r with respect to either of the above measures.

Many speakers focused on recent advances in interpolatory model reduction, in which the goal is to find a reduced order model with transfer function $\mathbf{H}_r(s)$ that tangentially interpolates the original transfer function $\mathbf{H}(s)$ at selected points $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ in selected tangential directions $\{\mathbf{c}_i\}_{i=1}^r \subset \mathbb{C}^p$ and $\{\mathbf{b}_j\}_{j=1}^r \subset \mathbb{C}^m$:

$$\mathbf{c}_i^T\mathbf{H}(\sigma_i) = \mathbf{c}_i^T\mathbf{H}_r(\sigma_i),\ \mathbf{H}(\sigma_j)\mathbf{b}_j = \mathbf{H}_r(\sigma_j)\mathbf{b}_j$$
and
$$\mathbf{c}_i^T\mathbf{H}_r'(\sigma_i)\,\mathbf{b}_i = \mathbf{c}_i^T\mathbf{H}'(\sigma_i)\,\mathbf{b}_i \text{ for } i = 1, \ldots, r.$$

Not surprisingly, a crucial issue in attempts to make $\mathcal{S}_r \approx \mathcal{S}$ via tangential interpolation lies in the choice of interpolation points and tangent directions. Interpolation conditions yielding *optimal* $\mathcal{H}_2$ system approximation are known and lead to a variety of computationally effective approaches. Kyle Gallivan (Florida State) presented new results along these lines, with special focus on the thorny case in which the reduced system has repeated poles. Zlatko Drmać (Zagreb) offered an analysis of the optimal $\mathcal{H}_2$ problem through study of a related fixed-point iteration and discussed some surprising periodicities in the iteration maps. Garret Flagg (Virginia Tech) focused on conditions for interpolation leading to optimal $\mathcal{H}_\infty$ system error and provided insight, informed by potential theory, as to how optimal $\mathcal{H}_2$ interpolation strategies might be extended to the $\mathcal{H}_\infty$ setting.

Various speakers expanded the range of applicability of linear model reduction. Ulrike Baur (Chemnitz) discussed ap-proaches for model reduction of linear parametrized systems in which the system data is parameter-dependent and the re-duced models have to be accurate over a wide range of parameter values. Motivated by similar problems, Chris Beattie (Virginia Tech) showed how to generate structure-preserving system interpolants when the original transfer function $\mathbf{H}(s)$ is presented in the form of a general co-prime factorization $\mathbf{H}(s) = \mathbf{C}(s)\mathbf{K}(s)^{-1}\mathbf{B}(s)$. Thanos Antoulas (Rice) presented data-driven methods for the construction of a reduced model directly from input/output observations without access to the original state-space system description. These developments were motivated in part by the availability of vector network analyzers, which are able to rapidly acquire empirical input/output characteristics in order to verify circuit models.

Timo Reis (TU Berlin) presented an ap-proach to passivity-preserving model reduction for electrical circuits described by a system of differential–algebraic equations (effectively resulting in a singular $\mathbf{E}$ matrix in the description of $\mathcal{S}$). Chris Massey (Army Corps of Engineers) described the application of interpolatory optimal $\mathcal{H}_2$ model reduction techniques in simulations of the linearized shallow water equations used to model the storm surge in Bay St. Louis (where Hurricane Katrina made landfall in 2005). Peter Benner (Chemnitz) illustrated the use of model reduction to solve linear inverse problems.

Large-scale *nonlinear dynamical systems* add a forbidding set of challenges to those encountered in model reduction of large-scale linear dynamical systems. With nonlinear reduced models of the form (1), a major performance bottleneck typically arises in the evaluation of the term $\mathbf{f}_r(\mathbf{y}_r,\mathbf{u})$. It is necessary to evaluate $\mathbf{f}$ on a "lifting" of $\mathbf{y}_r$ to the (large-scale) n-vector that it is in-tended to represent. Unless the nonlinear function $\mathbf{f}$ has a special form that allows precomputation of the lifting, evaluation, and compression process, naive evaluation of $\mathbf{f}_r(\mathbf{y}_r,\mathbf{u})$ will require at least $\mathcal{O}(n)$ function evaluations, independent of how small r may be. Depending on the choice of the outputs $\mathbf{\Theta}$, the same problem can also occur in the evaluation of $\mathbf{g}_r(\mathbf{y}_r,\mathbf{u})$.

Danny Sorensen (Rice) developed a solution to this severe bottleneck and presented the *discrete empirical interpolation*. Steve Cox (Rice) demonstrated the utility of this approach and introduced promising applications for model reduction in neurophysiology; Figure 1 shows the ability of a reduced model to approximate complex cellular neuronal dynamics.

Traian Iliescu and Jeff Borggaard (both of Virginia Tech) illustrated applications of model reduction in the simulation of turbulent flow. Karen Willcox (MIT) showed that model reduction can be extremely effective in uncertainty quantification and decision-making under uncertainty. Describing challenges encountered in complex chip design, Wil Schilders (NXP Semiconductors) proposed new approaches from graph theory for tackling these challenges. Similarly, Jacob White (MIT) introduced model reduction techniques and discussed their application in predicting cell-averaged behavior from biochemical kinetic models.

The speakers in the last session of the minisymposium related concepts from numerical linear algebra to model reduction. Daniel Kressner (ETH) discussed the use of Krylov subspace methods in solving large-scale matrix equations, such as Lyapunov equations, which commonly occur in the context of model reduction. Reduced order models do not always inherit stability from their full-order system origins, leaving open the question of how best to recover stability without destroying other qualities of the reduced system approximation. Mark Embree (Rice) approached this question, focusing on convergence and shift behavior for Arnoldi methods in the context of model reduction; as a by-product, he gave some early hints of an analog to eigenvalue interlacing for nonnormal matrices. In the final talk, Lizette Zietsman (Virginia Tech) connected the concept of distance to controllability with the actuator placement problem.

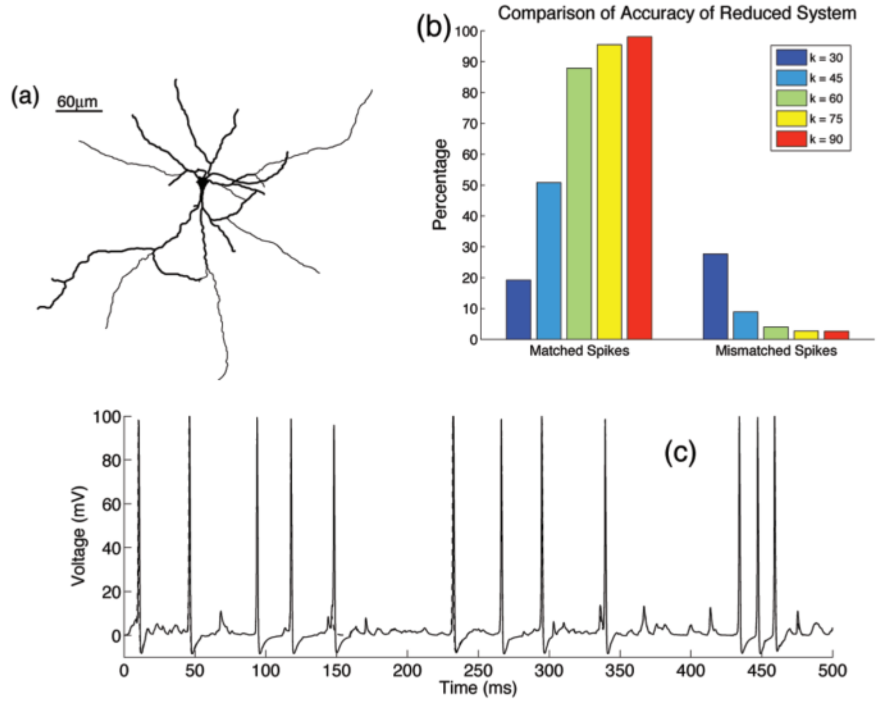

Figure 1. *Spike-capturing accuracy of the reduced versus the full system for the neuron shown in (a). The full system models the voltage response to random distributed synaptic input at 2233 points (one every micron). The same input is used to drive a drastically reduced system that captures the voltage at* k *points. The reduced system of order 75 is able to capture more than 90% of the true spikes (b), while generating very few false spikes. The full and reduced somatic spike trains over the initial half second of excitation are reproduced in (c). Courtesy of Steve Cox, Rice University.*

The model reduction minisymposium fielded a diverse group of outstanding speakers who together covered topics ranging from new theoretical aspects of the area to significant algorithmic advances to promising application directions. For those not working in the area, the minisymposium provided an overview of the area and a sense of state-of-the-art techniques (handy for those of us who do work in model reduction as well). Novices and experts left Miami energized, if not tanned, ready to take on some of the fresh model reduction challenges that have emerged—and with time to prepare for the next SIAM CSE meeting.

Christopher Beattie and Serkan Gugercin are in the Department of Mathematics at Virginia Tech.